Proposed Framework for Anomalous Change Detection

James Theiler & Simon Perkins
Space and Remote Sensing Sciences
Los Alamos National Laboratory

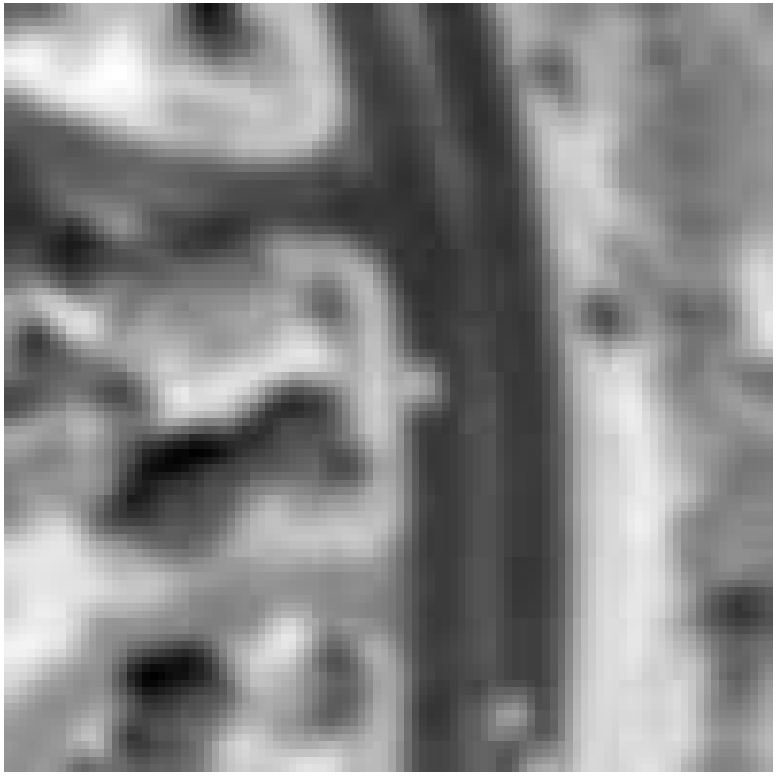
June 29, 2006

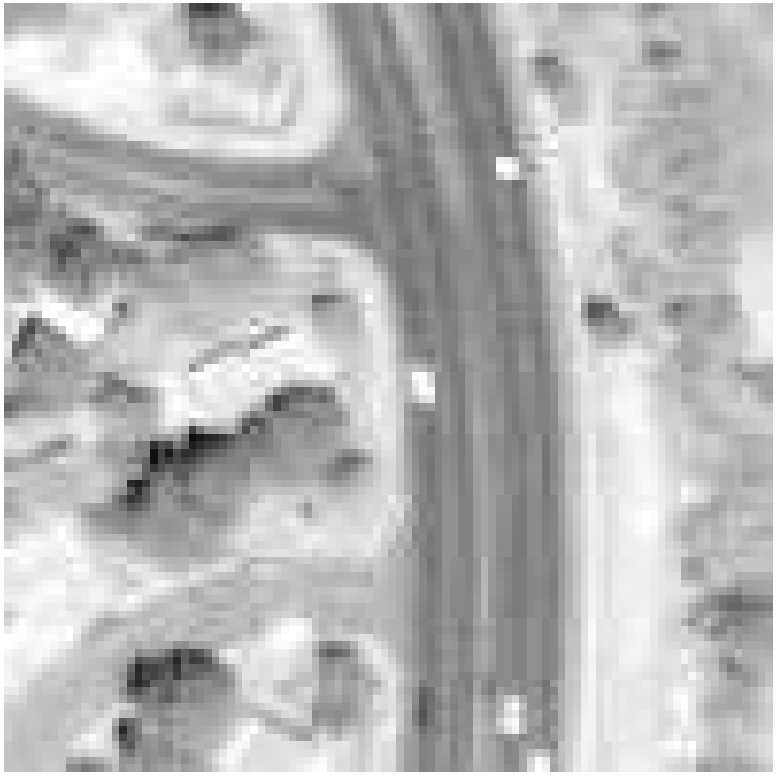
ICML Workshop on:
Machine Learning Algorithms for Surveillance and Event Detection

Motivation: what has changed?

- Illumination: Right image is brighter but has less contrast
- Focus: Right image is sharper
- Content?

What has changed? (closeup)





- Illumination: Right image is brighter but has less contrast
- Focus: Right image is sharper
- Content: There's an extra vehicle

Changes and Anomalous Changes

- Pixels from first image: $x_1, x_2, \ldots, x_n$
- Same pixels from the second image: $y_1, y_2, \ldots, y_n$
- Images in general are multi- (or even hyper-) spectral
  - But the two images need not be taken from the same camera, or even have the same number of spectral channels: $x_i \in \mathbb{R}^{d_x}$ and $y_i \in \mathbb{R}^{d_y}$
  - Even single-channel (panchromatic) imagery can be extended to "pseudomultispectral" using spatial operators

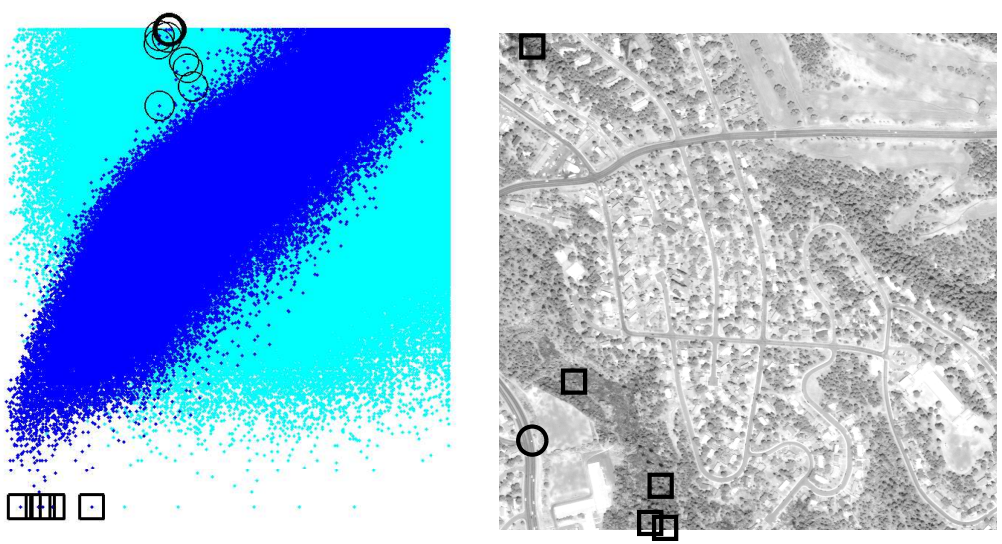
- The Change Detection Problem:
  - For what pixels i are x_i and y_i "different"?
  - Different? Well, "different" in an interesting way.
- The Anomaly Detection Approach:
  - Consider pixel pairs: $\{(x_1, y_1), (x_2, y_2), \ldots\}$
  - Assume (x_i, y_i) drawn from a distribution $P(x, y)$
  - Ask: for what pixels i are pairs (x_i, y_i) "unusual"?

Various approaches for Anomalous Change Detection

- **Straight anomaly detection**
  - Anomalies are rare: occur at low values of $P(x, y)$
- **Difference methods**
  - Simple difference: anomalous changes are large values of $|x - y|$
  - Regression: model $y \approx \hat{y} = \mathcal{L}(x)$, and look for large values of $|y - \hat{y}|$.
    - Or: model $x \approx \hat{x} = \mathcal{L}'(y)$, and look for large $|x - \hat{x}|$
  - Equalized difference:
    - Fit functions $\mathcal{H}_x(x)$ and $\mathcal{H}_y(y)$ to "equalize" x and y
    - Then look for large values of the difference $|\mathcal{H}_x(x) - \mathcal{H}_y(y)|$.
- **Proposed Framework**
  - Contours of $\dfrac{P(x, y)}{P(x)P(y)}$
  - Can be implemented as binary classification:
    - Data (x_i, y_i) samples from the "normal" class: $P(x, y)$
    - Resampled data (x_i, y_j) represents the background class: $P(x)P(y)$

What's wrong with straight anomaly detection?

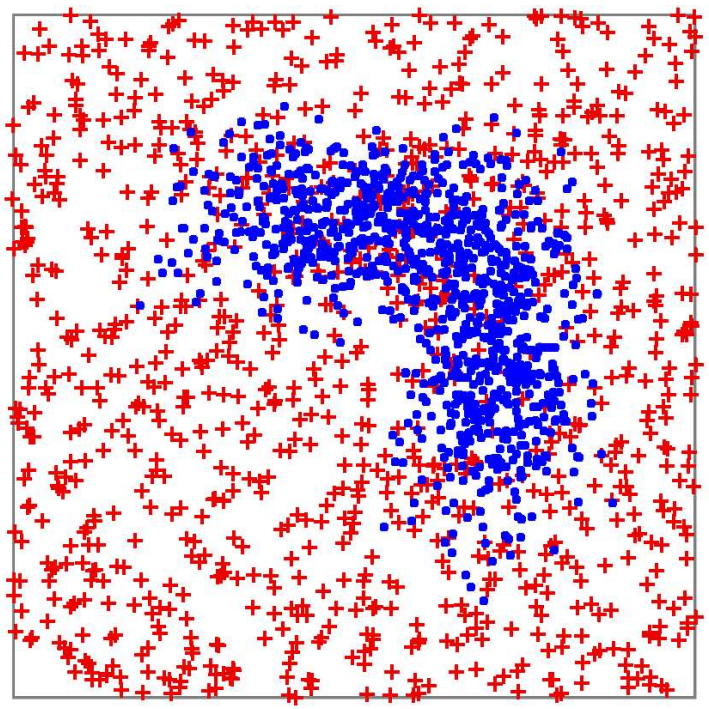
- Anomalies are rare: occur at low values of $P(x, y)$
- Straight anomaly detection schemes would consider particularly bright (or particularly dark) pixel pairs to be anomalous, even if there is no significant change.
- But what we mean (informally) by an "anomalous change" is
  - *not* that the components are individually unusual
  - but that the relationship *between* the components is unusual.



- Circles are "actual" changes
- Squares are pairs (x, y) where density $P(x, y)$ is low
  - Components are individually unusual; but
  - the $x \to y$ change is not so strange

What's so straight about straight anomaly detection?

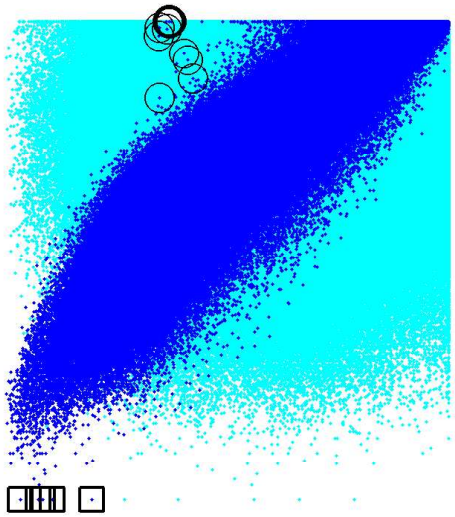
- Useful to recast anomaly detection as a binary classification problem
- Gory detail: see Steinwart *et al.* (2005).



- Uniform background transforms the anomaly detection problem into a binary classification problem.
  - "Normal" class is exemplified by the data (•)
  - Background class is defined by a uniform density, exemplified by a sampling from that density (+)
- Boundary separating normal from background produces a set of minimum area that encloses a maximum amount of data.

- So what's straight is that the background is uniform, but...
- With this recasting of the problem, nonuniform backgrounds are possible

Proposed framework uses nonuniform background

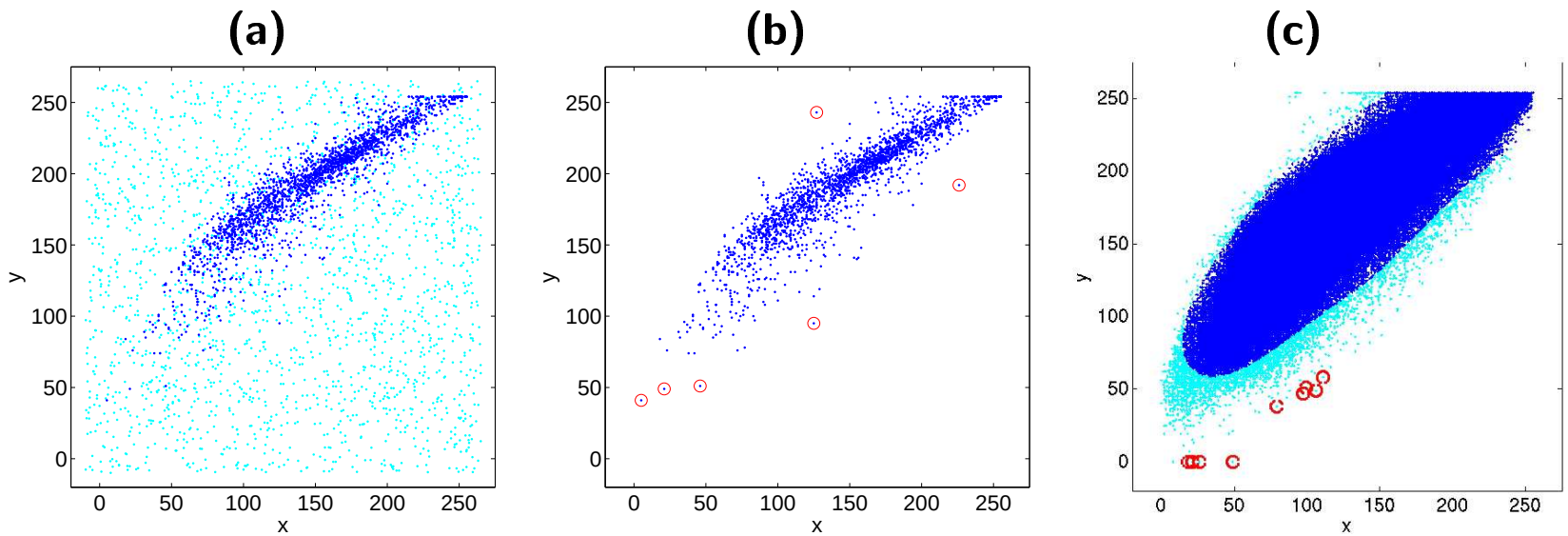


- Dark blue indicates data: $(x_i, y_i) \sim P(x, y)$
- Light blue indicates background: $(x_i, y_j) \sim P(x)P(y)$

- Anomalies of interest are in low-density regions of data: $P(x, y)$ small
- And in high-density regions of background: $P(x)P(y)$ large.
  - So circled points would be considered anomalous changes
  - but squared points would not be anomalous changes

- This distinguishes pixels which are anomalous in their own right (unusually bright or dark, for instance) from pixels in which the change from x to y is anomalous.

Use SVM for straight anomaly detection



(a) Train classifier using a sample of points

- Gaussian kernel, heavily weighted (10:1) normal class

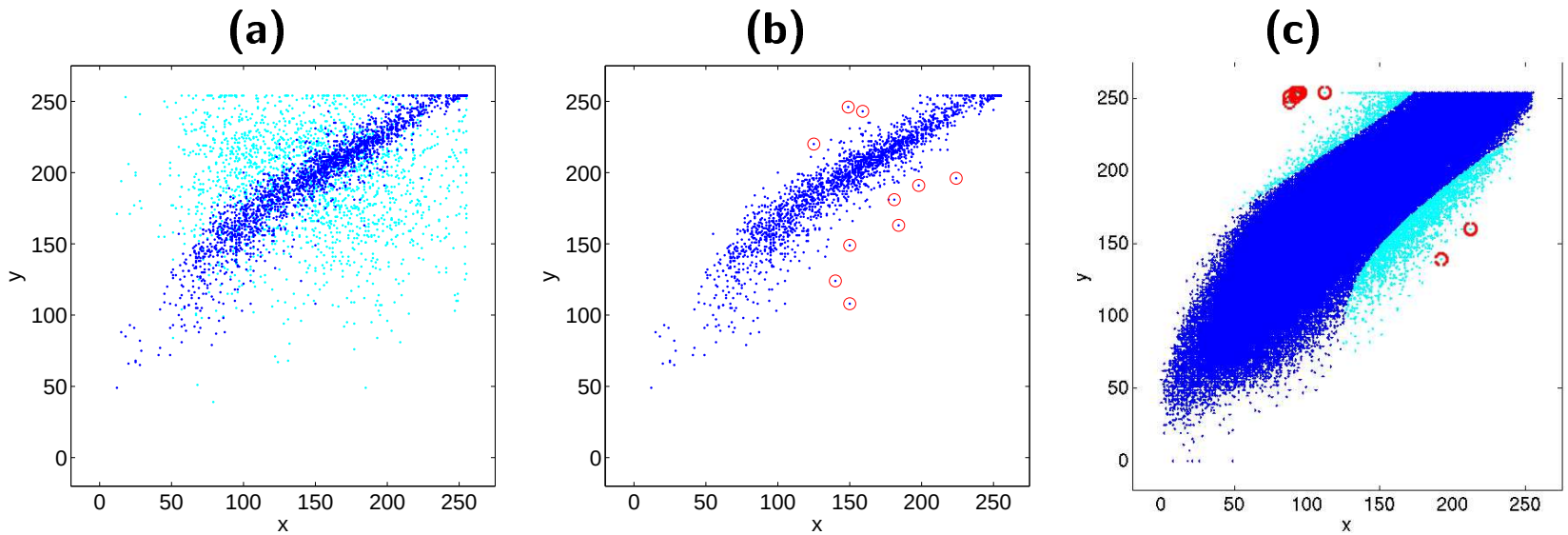
(b) Result of classifier applied to (normal) training data

- Red circles identified as anomalies

(c) Result applied to all normal data

- Cyan points identified in the anomalous class
- Red circles indicate the 9 most anomalous points

Use SVM to find anomalous changes (proposed framework)



(a) Train classifier using a sample of points

- Gaussian kernel, heavily weighted (10:1) normal class

(b) Result of classifier applied to (normal) training data

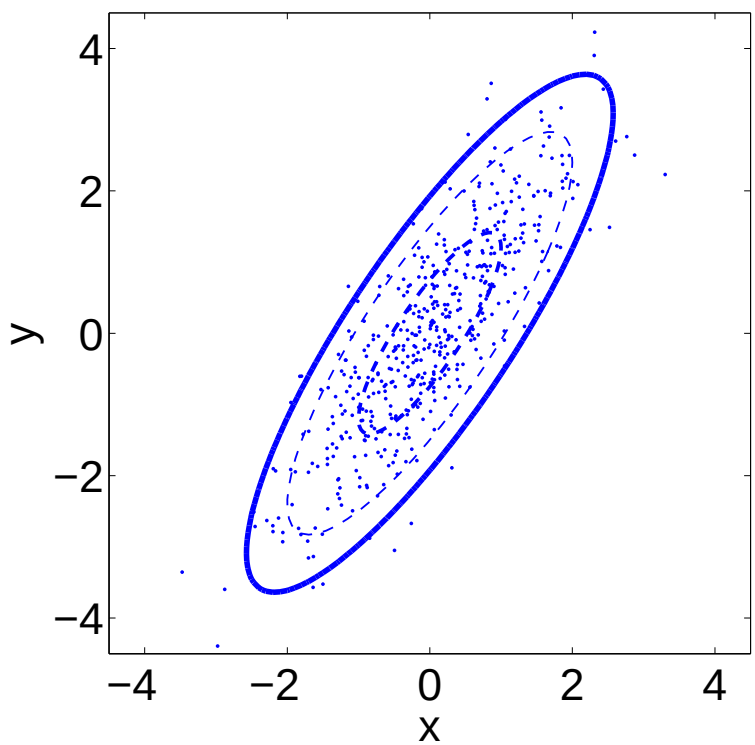
- Red circles identified as anomalous changes

(c) Result applied to all normal data

- Cyan points identified in the anomalous class
- Red circles indicate the 9 most anomalous changes

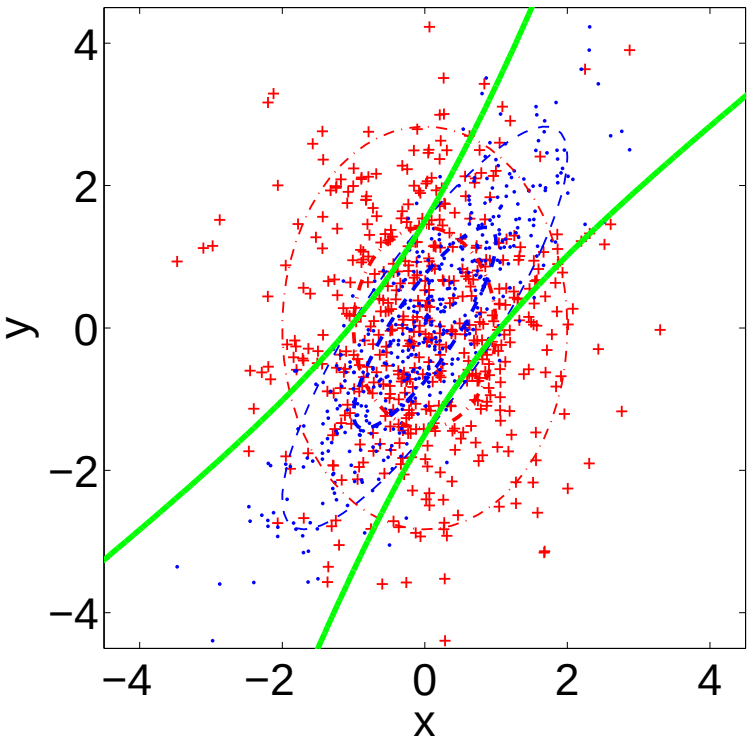
Gaussian example: straight anomaly detection

- Specified by covariances: $X = \langle xx^T \rangle$, $Y = \langle yy^T \rangle$, $C = \langle yx^T \rangle$.
- Level curves of $P(x, y)$ are contours of $\left[x^T \; y^T\right] K \begin{bmatrix} x \\ y \end{bmatrix}$,
  - where $K = \begin{bmatrix} X & C^T \\ C & Y \end{bmatrix}^{-1}$.
  - Contour is elliptical since K is positive definite.



- Here, $d_x = d_y = 1$
- $X = 1$, $Y = 2$, $C = 1.2$
- Dashed lines:
  - one and two sigmas
- Solid line:
  - encloses 95% of the data
  - minimum-area contour

Gaussian example: proposed framework



- Binary classification
  - • normal data (x_i, y_i)
  - + resampled background (x_i, y_j)
- Bayes-optimal classifier given by contour of the likelihood ratio:

$$\frac{P(x,y)}{P(x)P(y)}$$

- Solid lines enclose 95% of data

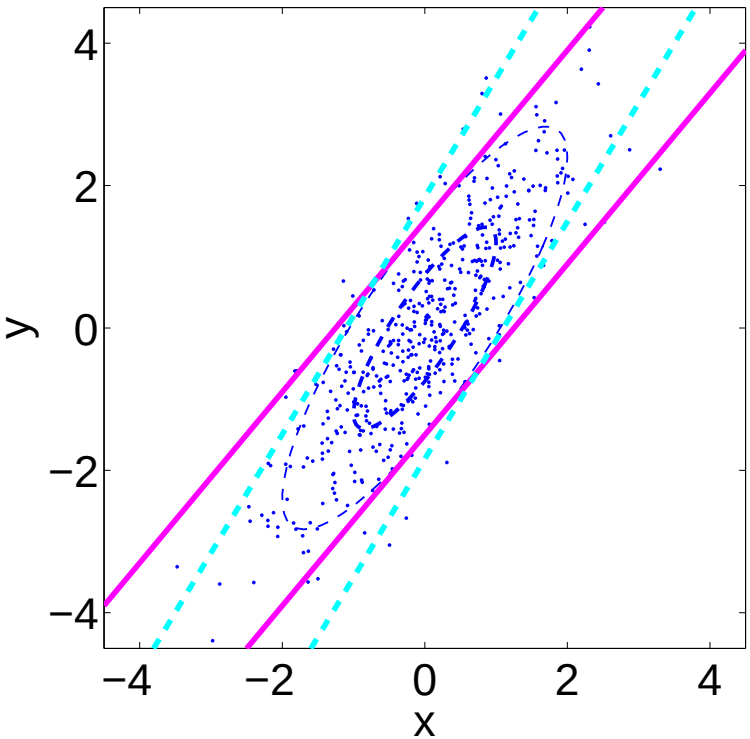
- Level curves of $P(x,y)$ are contours of $\begin{bmatrix} x^T & y^T \end{bmatrix} K \begin{bmatrix} x \\ y \end{bmatrix}$,
  - where $K = \begin{bmatrix} X & C^T \\ C & Y \end{bmatrix}^{-1} - \begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix}^{-1}$
  - Contours are hyperbolic since K is not positive definite.

Difference-based change detection: the "chronochrome"

- Developed for hyperspectral data (Schaum & Stocker, 1998)
- Fit linear*predictor: $y \approx \hat{y} = Lx$
  - Error matrix: $E = \left\langle (y - Lx)(y - Lx)^T \right\rangle$
  - Minimize trace of error matrix; obtain: $L = CX^{-1}$
- Vector-valued change: $\varepsilon = y - CX^{-1}x$
- Scalar anomalousness: $\boxed{e^2 = \varepsilon^T \left\langle \varepsilon\varepsilon^T \right\rangle^{-1} \varepsilon}$
  - Mahalanobis distance from centroid of change space
  - $(y - CX^{-1}x)^T(Y - CX^{-1}C^T)^{-1}(y - CX^{-1}x)$
- Note: asymmetry in x and y
  - Alternative: fit predictor $x \approx \hat{x} = L'y$
  - Leads to $\varepsilon = x - C^T Y^{-1} y$
  - Anomalousness: $(x - C^T Y^{-1} y)^T (X - C^T Y^{-1} C)^{-1} (x - C^T Y^{-1} y)$

*Or nonlinear predictor, eg with neural network (Clifton, 2003)

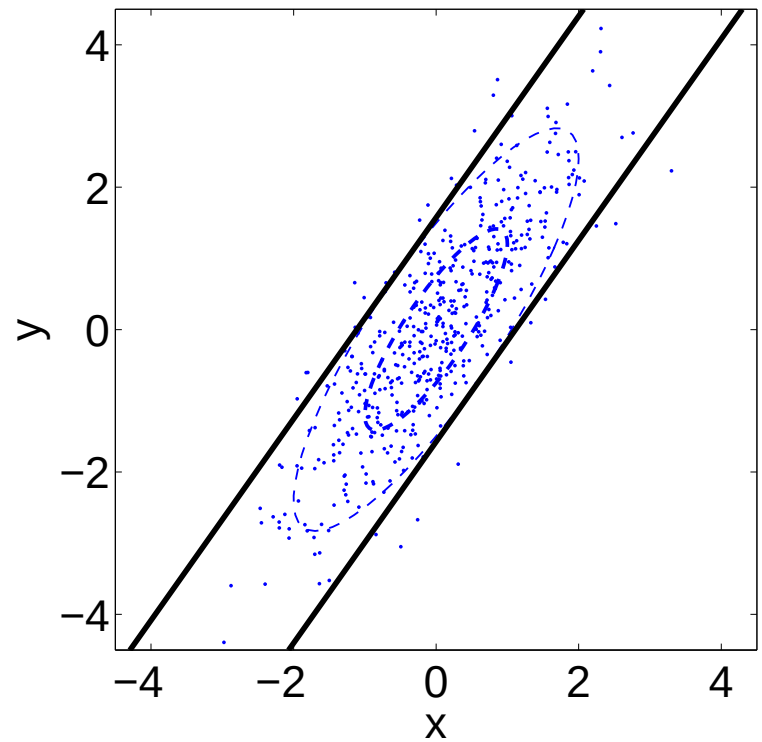
Gaussian example: the Chronochrome



- Linear contours
  - Cylindrical contours in higher dimensions
- Solid line: $|y - Lx|$
- Dashed line: $|x - L'y|$
- Note: asymmetry in x and y
  - How to choose?

Gaussian example: covariance equalization (CE)

- CE was introduced by Schaum & Stocker (1998) as an *approximation* to the chronochrome
- Uses $\hat{y} = L^* x$, with $L^* = Y^{1/2} R X^{-1/2}$, for "some" orthonormal matrix R
  - When $d_x = d_y$, Schaum & Stocker recommend $R = I$
  - equivalent formulation: $\varepsilon = X^{-1/2} x - Y^{-1/2} y$
  - Contours of $e^2 = \varepsilon^T \left\langle \varepsilon \varepsilon^T \right\rangle^{-1} \varepsilon$



- Unlike the chronochrome, CE...
  - is symmetric in x and y
  - gives "min-volume" linear solution
  - lacks solid mathematical basis

Anomalous Change Detection: a three-band example

- Two images of the same scene

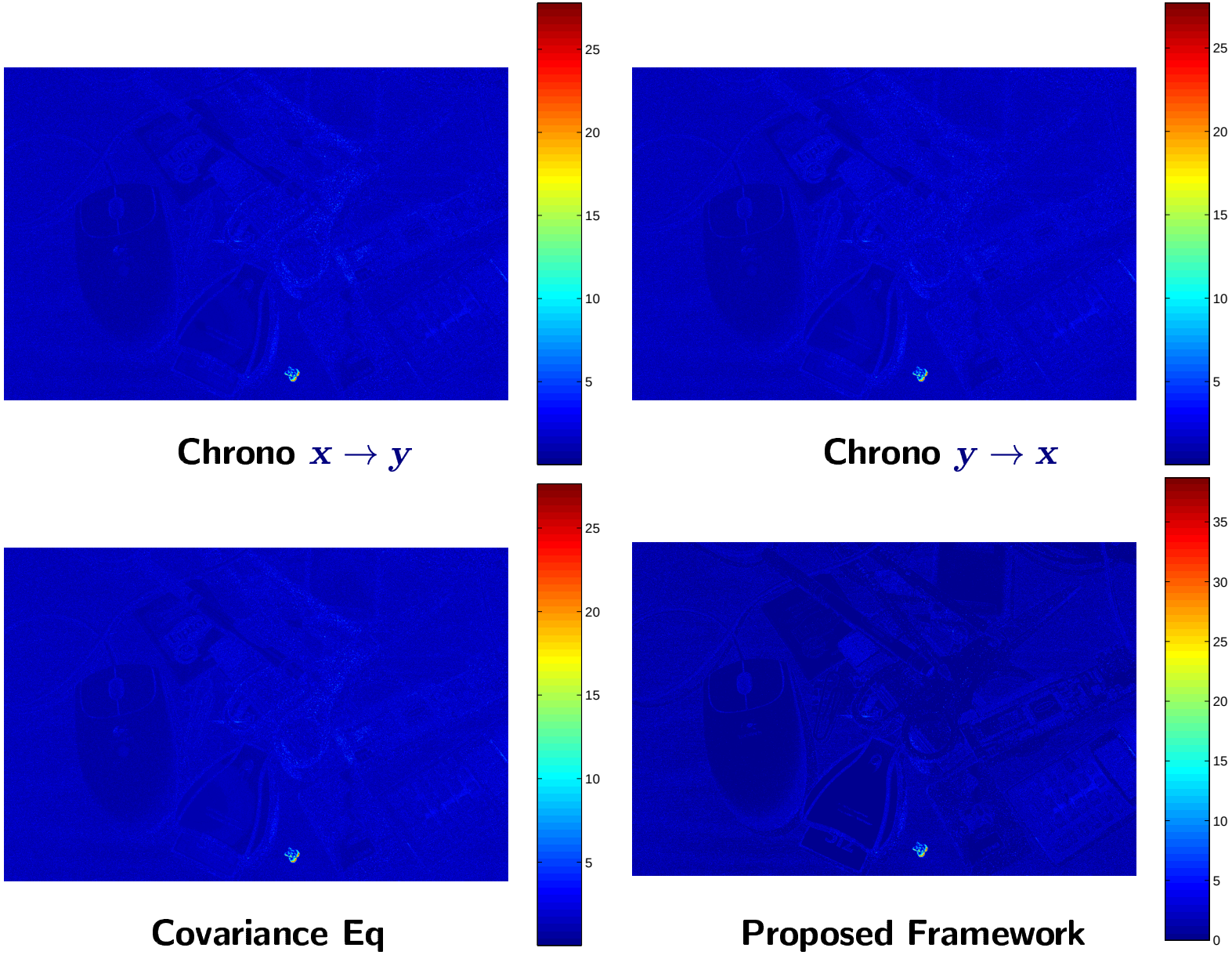
  - same camera, but different white-balance settings
  - tripod used to ensure good pixel co-registration
  - one small difference in content

Anomalous Change Detection: a three-band example

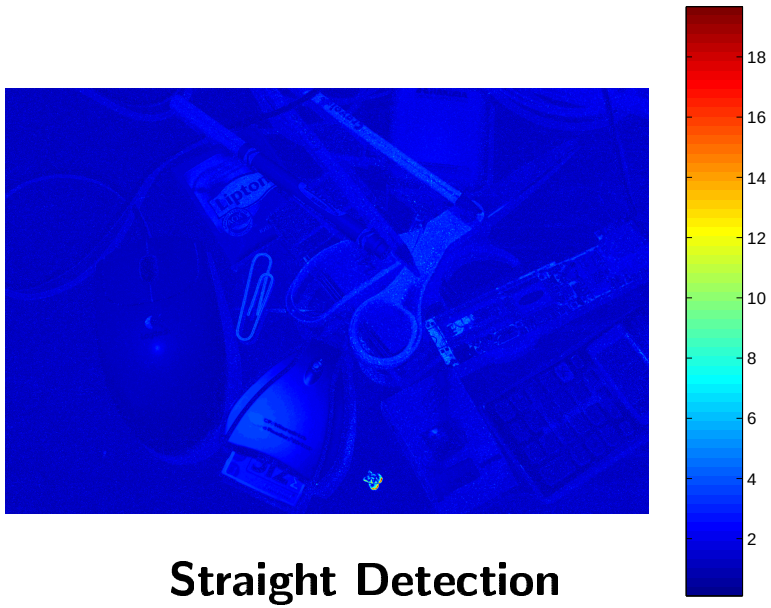
- Two images of the same scene

  - same camera, but different white-balance settings
  - tripod used to ensure good pixel co-registration
  - one small difference in content

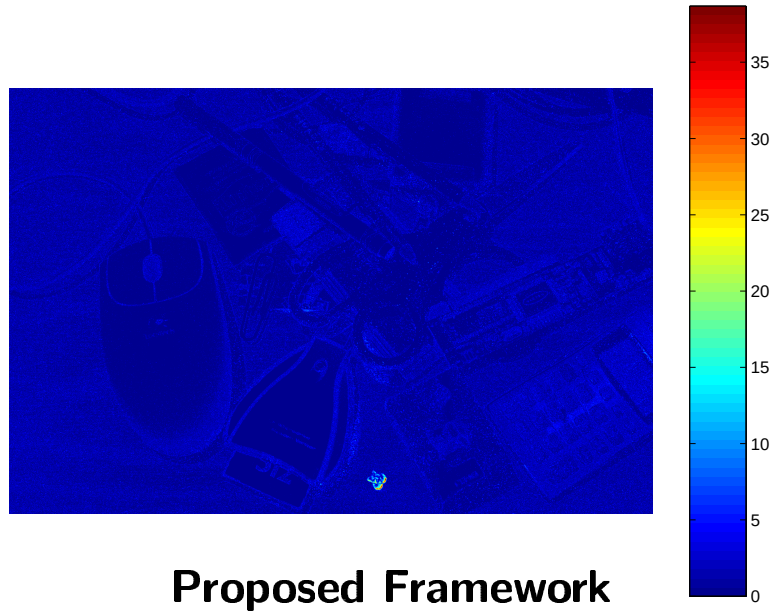
Comparison with anomalous change detectors



Comparison with straight anomaly detection



Straight Detection



Proposed Framework

- Let $s = \left[\mathbf{x}^T \; \mathbf{y}^T\right] \boldsymbol{K} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}$; image shows $e = \sqrt{s_+}$, scaled so $\left\langle e^2 \right\rangle = 3$
  - Scaling "equivalent" to chronochrome and covariance equalization
  - Here, $s_+ = s$ when $s \geq 0$; and $s_+ = 0$ when $s < 0$.
- Straight: $\boldsymbol{K} = \begin{bmatrix} \boldsymbol{X} & \boldsymbol{C}^T \\ \boldsymbol{C} & \boldsymbol{Y} \end{bmatrix}^{-1}$; Proposed: $\boldsymbol{K} = \begin{bmatrix} \boldsymbol{X} & \boldsymbol{C}^T \\ \boldsymbol{C} & \boldsymbol{Y} \end{bmatrix}^{-1} - \begin{bmatrix} \boldsymbol{X} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Y} \end{bmatrix}^{-1}$

Experiments with missing bands

(a) no blue

(b) no red

- Image (a) has only the Red and Green bands
- Image (b) has only the Green and Blue bands
- Both are two-band images ($d_x = d_y = 2$) but bands are not the same
  - this may be problematic for Covariance Equalization

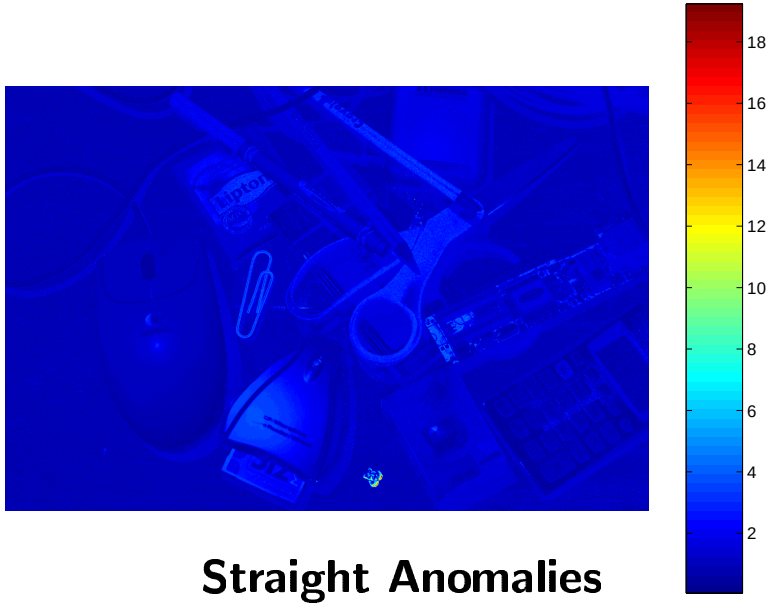
Experiments with missing bands

(b) no red

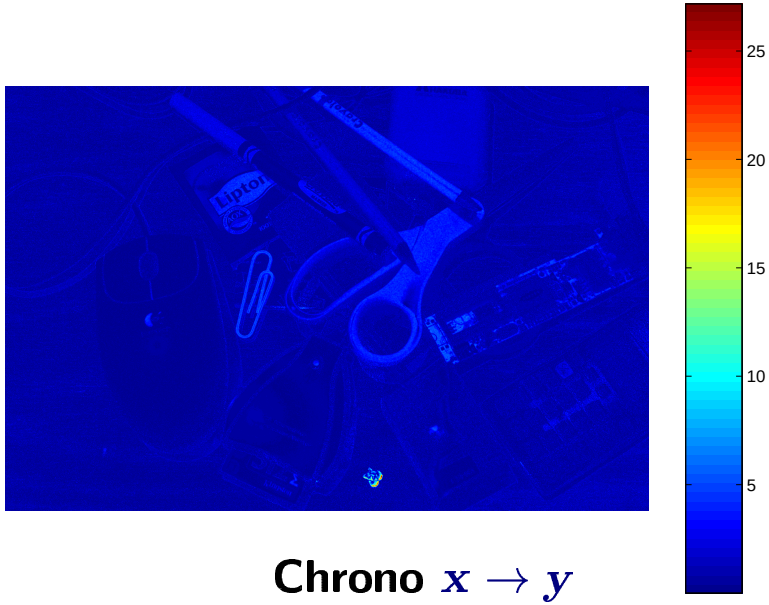
(a) no blue

- Image (a) has only the Red and Green bands
- Image (b) has only the Green and Blue bands
- Both are two-band images ($d_x = d_y = 2$) but bands are not the same
  - this may be problematic for Covariance Equalization

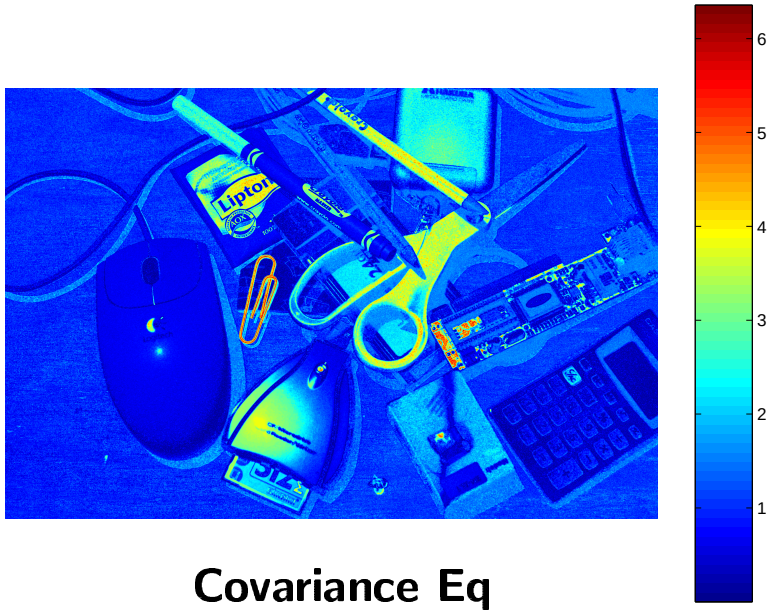
Missing bands: comparison of anomalous change detectors



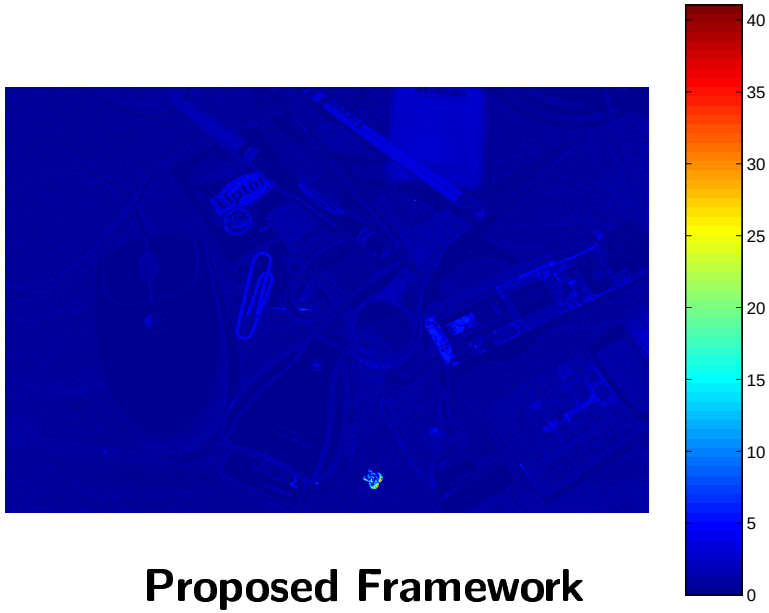
Straight Anomalies



Chrono $x \to y$



Covariance Eq



Proposed Framework

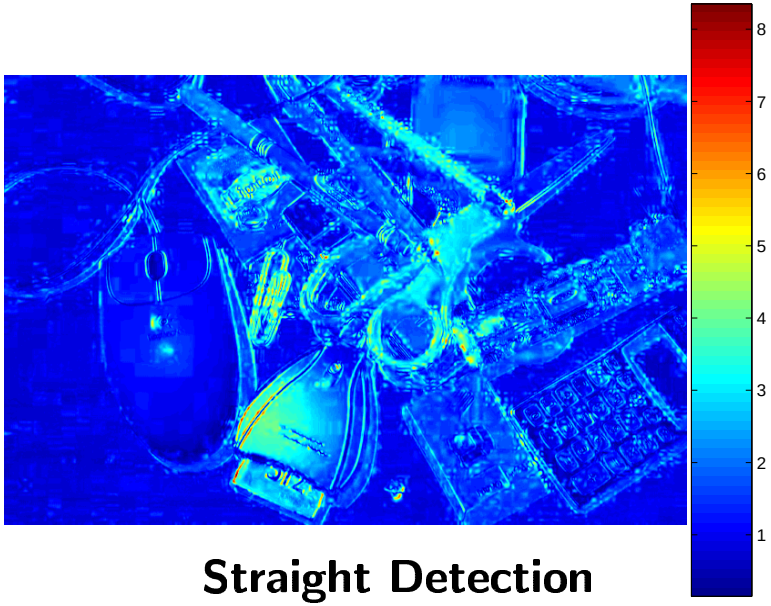
Experiments with degraded images

- Subsample from 2000×3008 to 500×752 pixels
- Misregister by one horizontal pixel in the subsampled image
- Blur image with a 3×3 kernel
- Re-crop to 500×700 pixels
- Now, look for anomalous changes

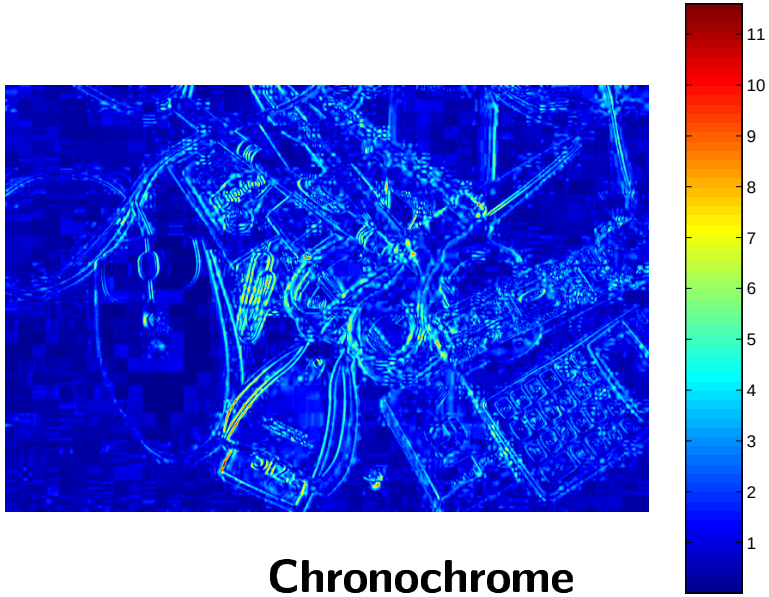
Experiments with degraded images

- Subsample from 2000×3008 to 500×752 pixels
- Misregister by one horizontal pixel in the subsampled image
- Blur image with a 3×3 kernel
- Re-crop to 500×700 pixels
- Now, look for anomalous changes

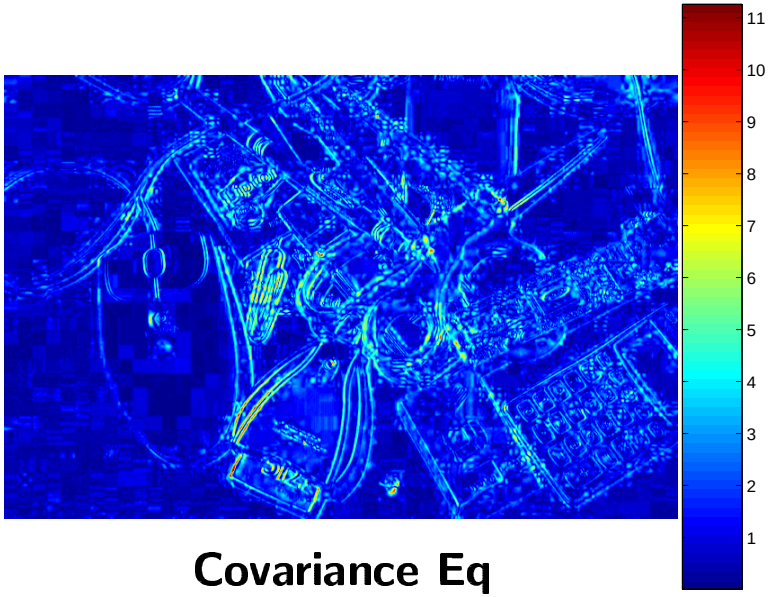
Degraded image: anomaly detectors



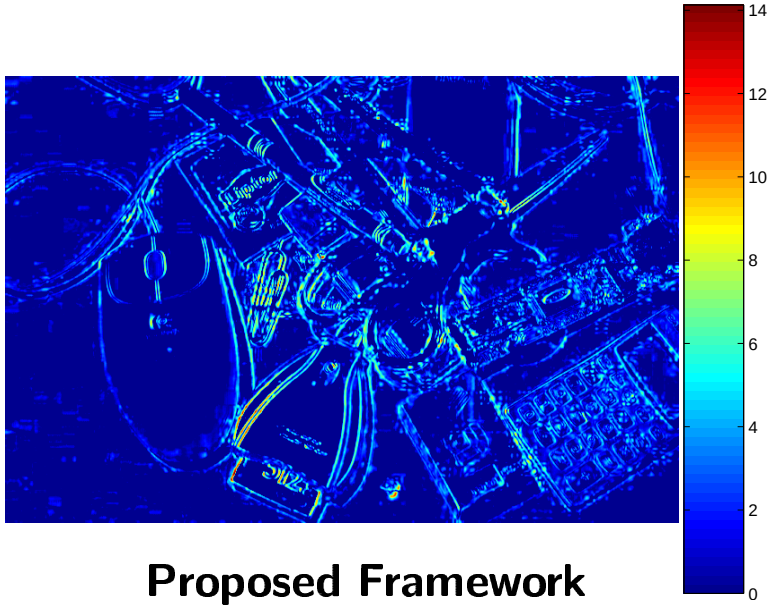
Straight Detection



Chronochrome

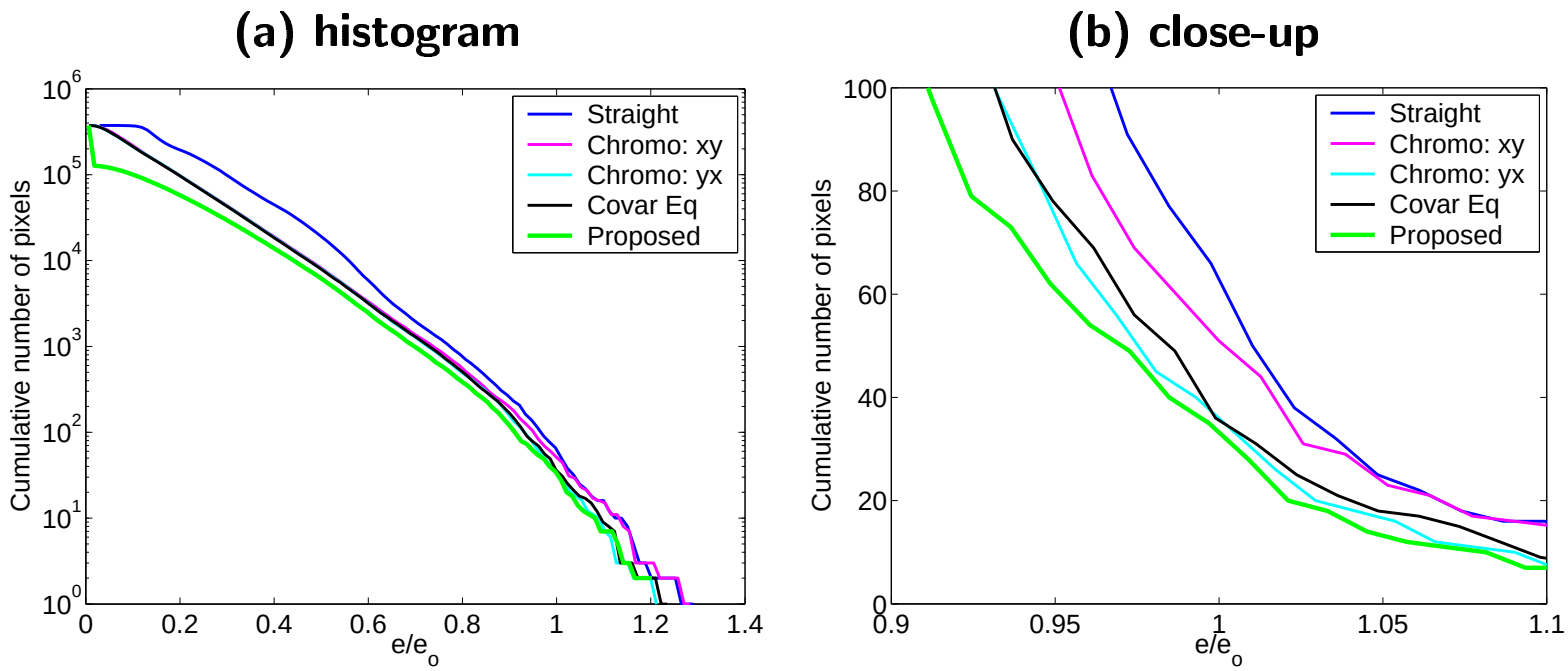


Covariance Eq



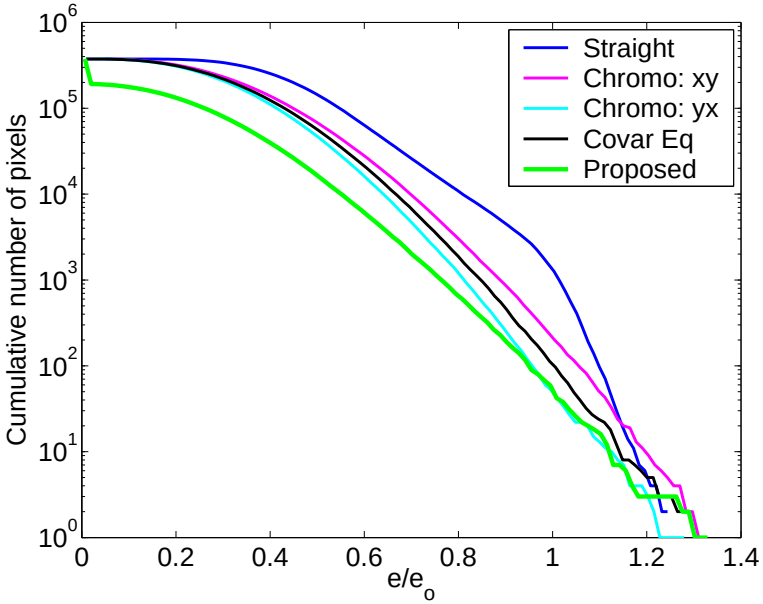
Proposed Framework

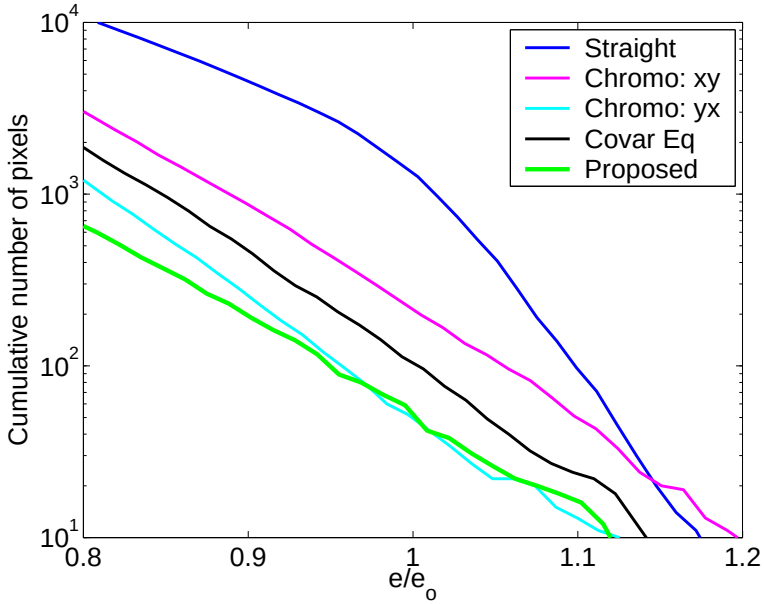
Degraded image: histograms of anomaly detection



- Cumulative histograms of e values in anomaly detection images
- Set e_o as the value associated with the actual change
- Pixels with $e/e_o > 1$ are false alarms
- Lower curves are better

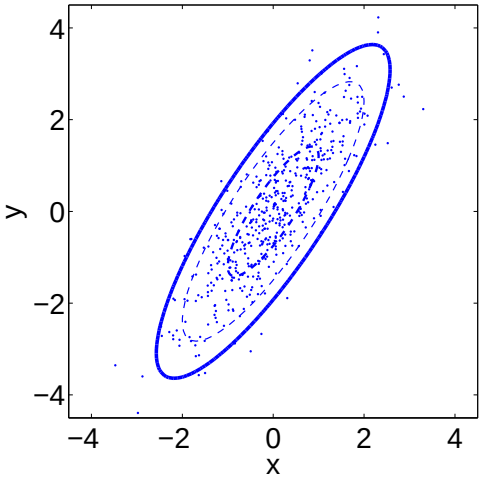
Further degraded image: add some noise



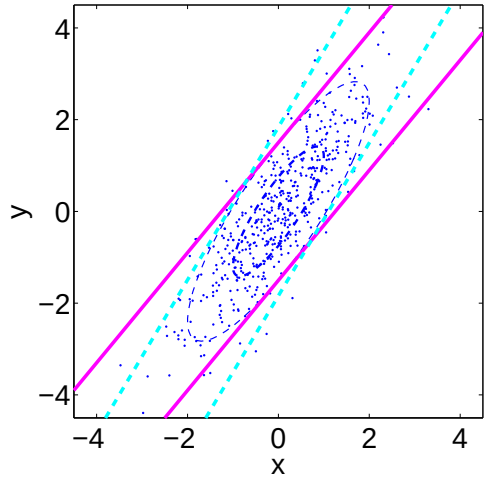


Summary of anomalous change detection frameworks

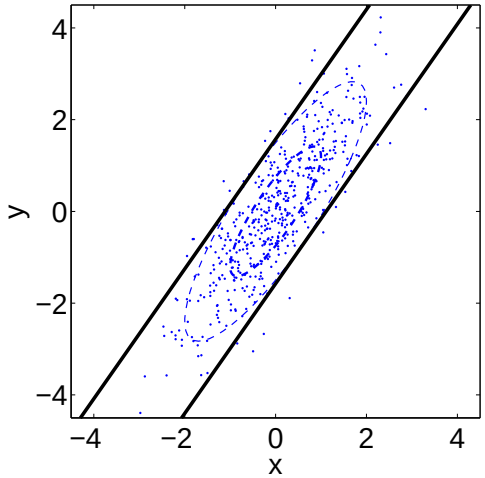
Straight Anomalies



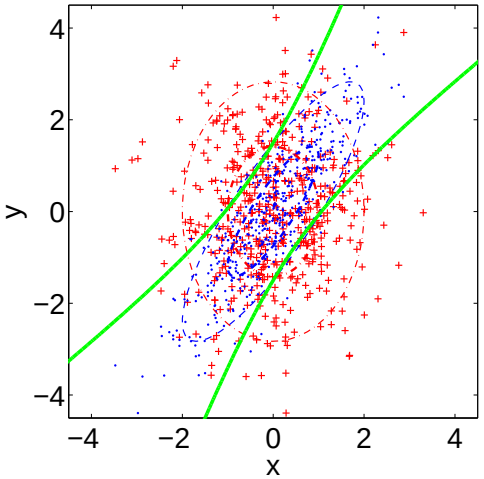
Chronochrome



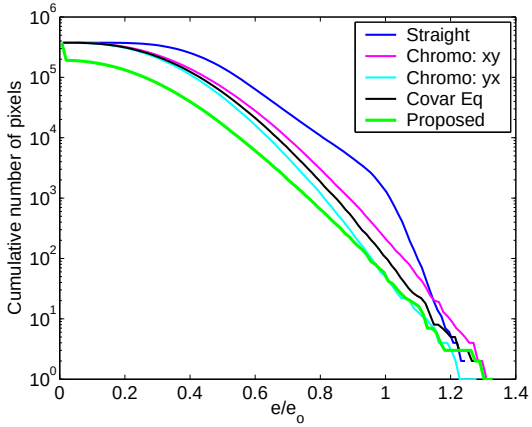
Covariance Eq

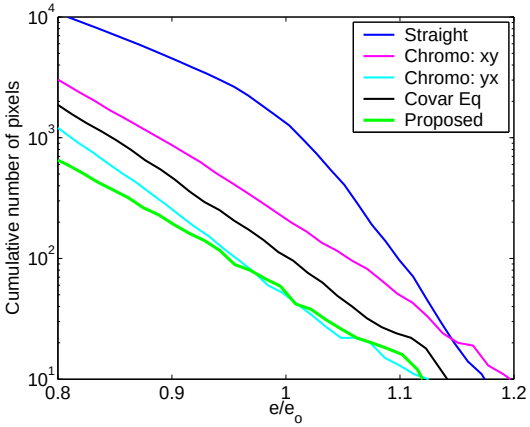


Proposed Framework

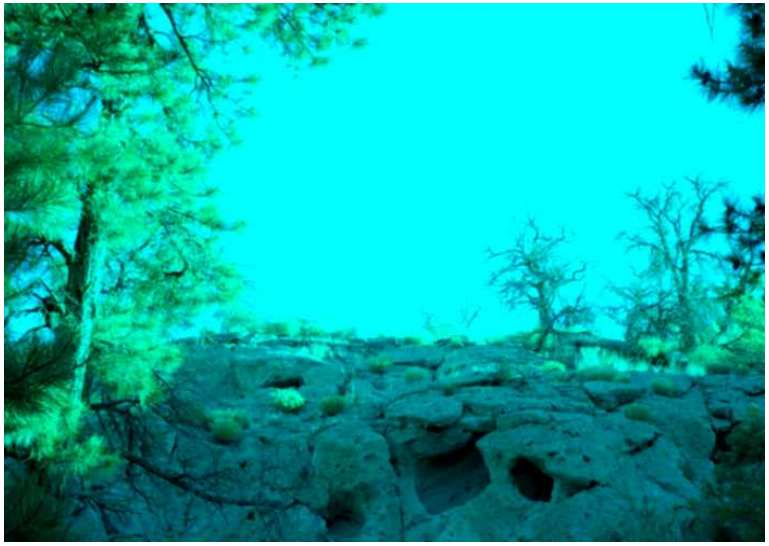


Cumulative Histograms





(Yet) another example: field experiment

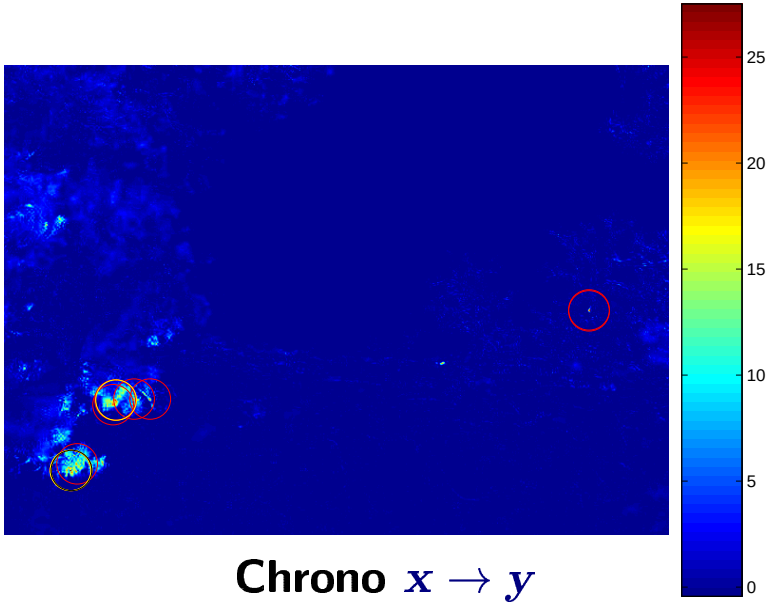


x

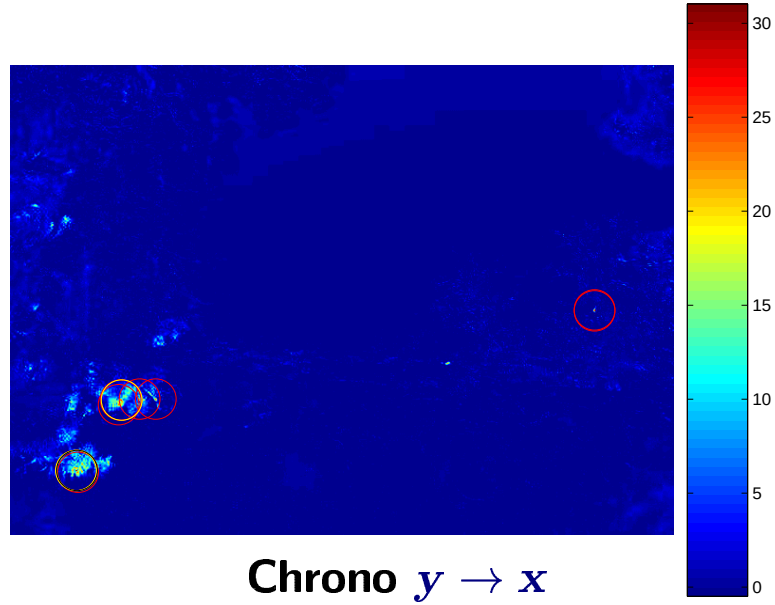
y

- Two frames taken several minutes apart, just after sunrise
  - Can see changes in lighting
  - There is also a bird in the tree (in one of the frames)
- Images downsampled to 500x752, then cropped to 500x700
- Red/Blue channel removed from first/second image

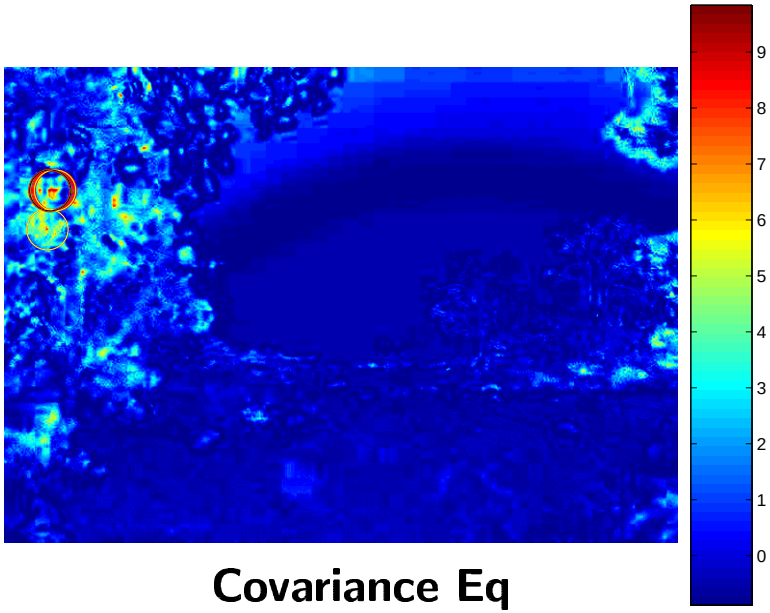
Field Experiment



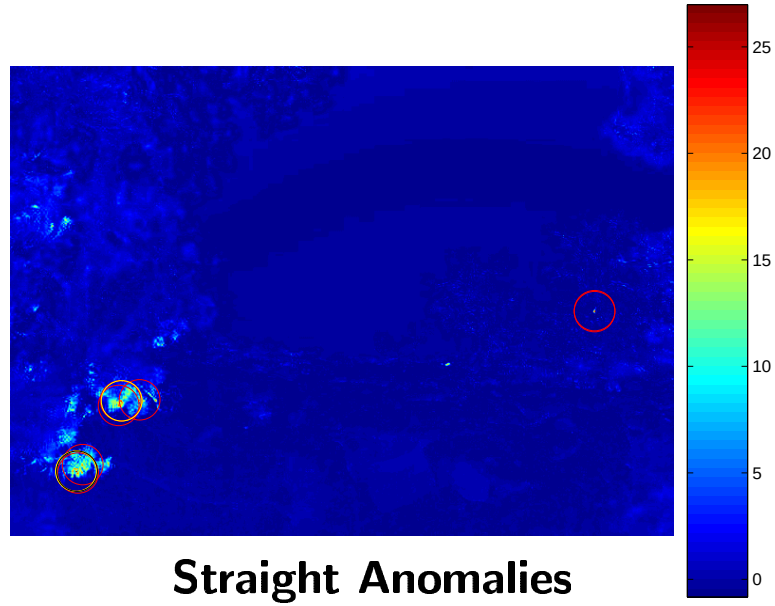
Chrono $x \to y$



Chrono $y \to x$

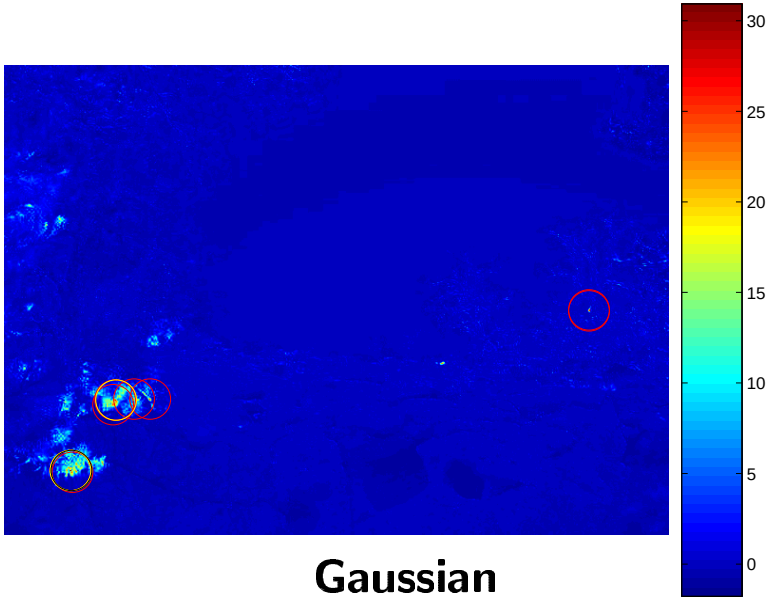


Covariance Eq

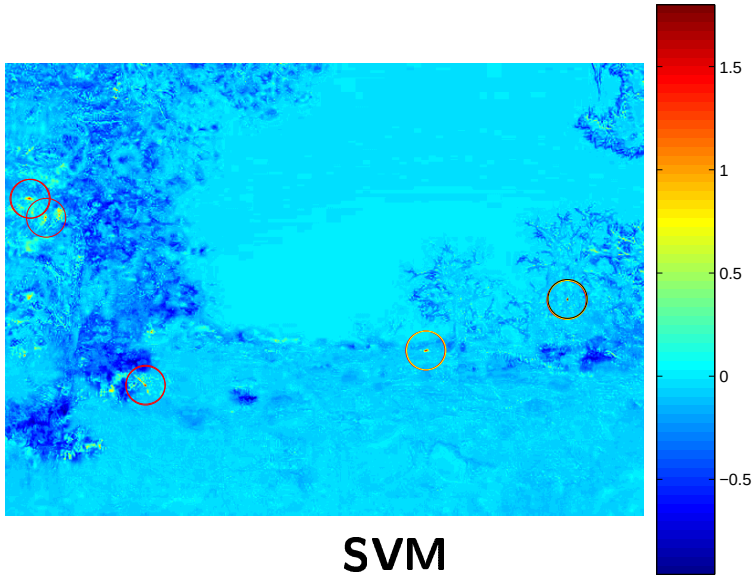


Straight Anomalies

Field Experiment: using proposed framework



Gaussian

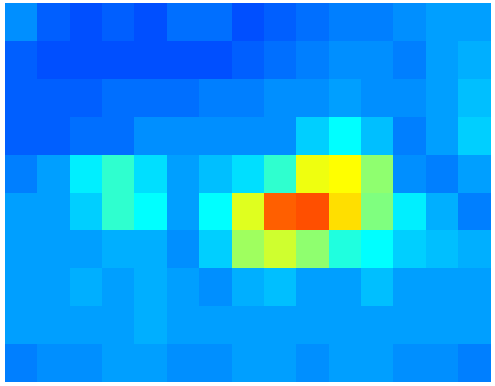


SVM

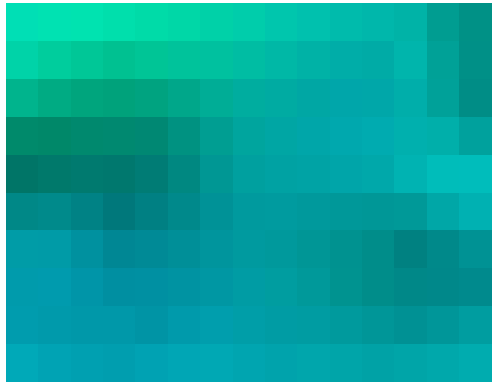
- Ten most anomalous changes identified with circles
  - (7)Red→(2)Yellow→(1)Black: increasing anomalousness
- Gaussian: Bird in tree is a red-circle anomaly
- SVM: Bird in tree is the most anomalous (black circle) change
  - Pixels near the bird have a yellow and a red circle
- SVM: New anomalous change (yellow circle) identified below the tree

Field Experiment: follow up on a new anomalous change

- **Close-up on anomalous change:**



:



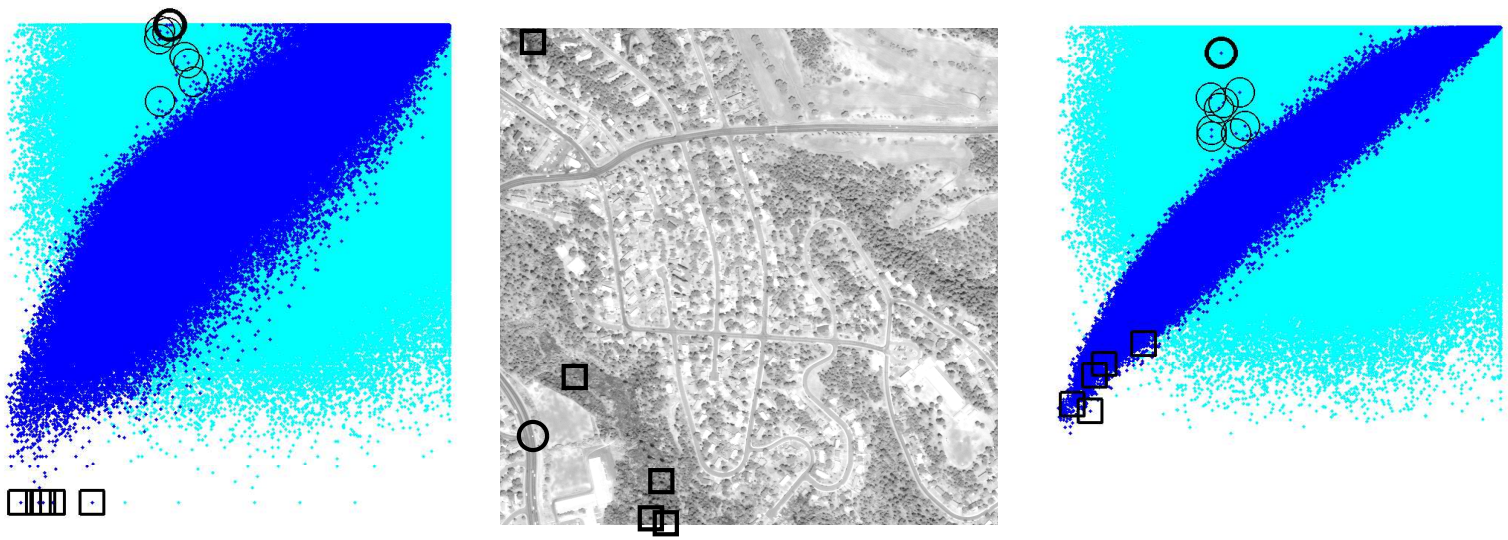
$\rightarrow$

Anomalousness **x** **y**

- **Close-up on original (un-degraded) images:**

$\rightarrow$

Issues that arise specifically with imagery

The formulation of anomalous change detection does not explicitly require that data be organized as rectangular arrays of pixels, but imagery provides both a natural setting and a compelling application for the change detection problem.

- **Application of spatial operators: dilation, erosion, smoothing, etc.**
  - **Can help with misregistration and with noise reduction**
  - **Can help "equalize" images (as seen below)**
  - **Leads to question: how to learn best spatial operations?**
    - **Can generate multiple bands – a problem in feature selection**